Exhibit 6

JOURNAL OF COMPUTATIONAL PHYSICS 61, 463-482 (1985)

Solving or Resolving Inadequate and Noisy Tomographic Systems

GUUST NOLET

Department of Theoretical Geophysics,
P.O. Box 80.021, 3508 TA Utrecht, The Netherlands

Received June 27, 1984

Tomography in seismology often leads to underdetermined and inconsistent systems of linear equations. When solving, care must be taken to keep the propagation of data errors under control.

In this paper I test the applicability of three types of damped least-squares algorithms to the kind of sparse matrices encountered in seismic tomography: (1) singular value decomposition with Lanczos iteration, (2) conjugate gradient iteration with the LSQR algorithm, and (3) the Dines–Lytle method.

Lanczos iteration may be applied to large sparse systems of low rank to calculate solutions by singular value decomposition but becomes impractical with problems of larger size. The Paige–Saunders algorithm (LSQR), which incorporates Lanczos' iteration into a conjugate gradient method, provides a least-squares solution to the system with acceptable filtering properties. In a synthetic tomographic experiment, it proved to be converging up to an order of magnitude faster than the Dines–Lytle algorithm, a stationary iterative process.

For large tomographic systems, where restrictions in the available computer time pose limitations on the number of iterations, this indicates that conjugate directions methods are to be preferred to the more commonly applied Gauss–Seidel type of algorithms.

To avoid unwarranted conclusions when the problem is severely underdetermined or undermined with large data errors, resolution analysis must be applied to the data set. An algorithm for the determination of the resolution in sparse systems is given.

1. INTRODUCTION

As with other tomographic measurements, seismic data represent an integral property of the medium crossed by the ray path. Usually (but not always) the seismologist is interested in the seismic *velocity* within the medium, not in its absorbing properties. However, the mathematical problem is not different from other projection problems in electron microscopy or radiology.

Whether earthquakes or artificial sources are used as radiators of energy, the pattern of projection angles is generally inadequate in seismic tomography. The resolution provided by the system of crossing rays may differ substantially for different localities in the subsurface. The algebraic system is almost always ill-conditioned. Moreover, data errors can be large.

0021-9991/85 $3.00

In this paper I shall briefly compare several iterative matrix techniques for the calculation of the damped least-squares solution (DLS) in a synthetic tomographic experiment, and develop an efficient method to calculate the resolution of a given data set.

The three DLS methods explored in this paper are representative for three broad classes of inversion algorithms:

(1) Singular value decomposition: the Paige-Lanczos algorithm;

(2) Conjugate directions methods: the Paige Saunders algorithm;

(3) Stationary iterative processes: the Dines-Lytle algorithm.

Although seismology will act as *leitmotiv* throughout this paper, the conclusions and the methods to calculate the resolution are useful for any tomographic system with imperfect illumination.

2. Seismic Tomography

Tomographic methods have first been used in global seismology by Aki et. al. [1] and the number of applications in this field is growing fast (see Whitcombe [17] for further references). Some recent applications in exploration seismics are listed in McMechan [9].

In general we are interested in the seismic velocity $v(\mathbf{r})$ at location $\mathbf{r}$ within a region R. We shall denote the inverse of the velocity (the slowness) by $s(\mathbf{r}) = 1/v(\mathbf{r})$. The problem may be formulated in any number of dimensions (D) but usually $D = 3$. Sometimes shooting is done along one line only, so that the subsurface must be assumed homogeneous in one of the spatial coordinates and $D = 2$. On the other hand, time may be an independent parameter as well, e.g., when we monitor an oil reservoir or the underground gasification of a coal seam. In that case the theory in this paper can easily be adapted to include the time, if we scale time by some velocity v_0, appropriate for the velocity with which the model geometry is changing. In that case we set $D = 4$ and $r_4 = v_0 t$.

Because perturbations in the velocity field cause ray bending, the relationship between the travel times of seismic rays and $s(\mathbf{r})$ is nonlinear. We linearize the problem to first order by specifying a starting model $s_0(\mathbf{r})$. Deviations $s(\mathbf{r})$ to the starting model are related in an approximate way to deviations t_j' from the predicted travel time for the jth ray:

$$t_j' = \int_{S_j \cap R} s(\mathbf{r})\, d^D\mathbf{r} \qquad (j = 1,\ldots, N) \tag{1}$$

where $s_j \cap R$ is the intersection of the jth ray through R. In classical ray theory this intersection is assumed to be infinitely narrow, reducing (1) to a line integral. Fermat's Principle may be invoked to argue that the value of the integrand in (1) is

insensitive to small perturbations in S_j. We can therefore safely calculate S_j for the ray in the (known starting model. This is instrumental in the linearization of the problem.

In order to solve (1) we discretize the space R. We do this by defining a finite number of basis functions $h_k(\mathbf{r})$ that span a subspace of the Hilbert space of all slowness models $\{s(\mathbf{r})\}$:

$$s(\mathbf{r}) = \sum_{k=1}^{M} s_k h_k(\mathbf{r}). \tag{2}$$

Without loss of generality we may assume that the basis functions are orthonormal [10]:

$$\int_R h_i(\mathbf{r})\, h_j(\mathbf{r})\, d^D\mathbf{r} = \delta_{ij}. \tag{3}$$

In general we shall subdivide R into a number of nonoverlapping cells, so that we can satisfy (3) by a simple scaling of the basis functions. In the following we adopt

$$\begin{aligned} h_i(\mathbf{r}) &= v_i^{-1/2} \qquad \text{within cell } i \\ &= 0 \qquad\qquad \text{outside} \end{aligned} \tag{4}$$

where

$$v_i = \int_{\text{cell } i} d^D\mathbf{r}$$

is the cell's "volume." With this (2) reduces to a sparse system of linear equations

$$t'_j = \sum_k L'_{jk} s_k \tag{5}$$

where

$$L'_{jk} = \int_{S_j \cap R} h_k(\mathbf{r})\, d^D\mathbf{r}. \tag{6}$$

It will be convenient to scale the data to unit variance. If we denote the covariance matrix of the data by $\mathbf{C}_d$, this can be achieved by setting $\mathbf{t} = \mathbf{C}_d^{-1/2}\mathbf{t}'$, $\mathbf{L} = \mathbf{C}_d^{-1/2}\mathbf{L}'$, so that

$$\mathbf{Ls} = \mathbf{t}. \tag{7}$$

$\mathbf{L}'$ is a sparse matrix. This property will be destroyed in $\mathbf{L}$ by the scaling to unit variance, except if $\mathbf{C}_d^{-1/2}$ is a narrow-band matrix. Usually we lack information on the covariance of the data errors, and set $\mathbf{C}_d^{1/2} = \text{diag}[c_1, c_2, \ldots, c_N]$, with c_i the estimated standard deviation in delay time t_i.

Equation (7) is the equation we wish to solve in seismic tomography. In most recent applications the number of cells has not been larger than about 200. Advances in instrumentation and data transmission techniques may soon increase this figure by several orders of magnitude. A typical application in the near future may involve the inversion of, say, 200,000 arrival times for a model with $50 \times 50 \times 20 = 50{,}000$ cells, or a matrix with 10^{10} elements, of which about 10^7 are nonzero. Clayton [4] reported preliminary results of a tomographic interpretation of 1.7×10^6 data. This clearly asks for algorithms that avoid explicit storage of the elements of **L**. Such algorithms are invariably iterative. It is well known that their rate of convergence depends strongly on the type of matrix involved. Note that for very large systems it may be prohibitively expensive to do more than one iteration (at least with today's computers), even though such simple inversion ("backprojection") has serious deficiencies in the case of inadequate illumination. This paper assumes that it is feasible to do at least a few iterations.

The matrix **L** has an effective rank K which is in general much less than $\text{Min}(M, N)$, where M is the number of cells ("unknowns"). Because of data errors, nonlinear effects, or approximations in the calculation of the ray path geometry [15], the equations in (7) are not only underdetermined, but inconsistent as well. A Gauss transform reduces (7) to a consistent system of M equations

$$\mathbf{L}^T\mathbf{L}\mathbf{s} = \mathbf{L}^T\mathbf{t} \tag{8}$$

which leads to the least-squares solution of (7). If the elements of **L** were uncorrelated and randomly distributed around O, the left Gauss transformation would also have the effect of making $\mathbf{L}^T\mathbf{L}$ a matrix with a dominant diagonal. With tomography, however, all elements of **L** are nonnegative, and preferent ray geometries may introduce strong correlations between its columns. For this reason we do not expect that iterative methods of the Gauss–Seidel type will operate very effectively on this system.

In the following we shall study iterative solutions to (8). Although we shall often use the matrix notation, it is important to realize that none of the techniques described requires the matrix **L** or its transpose to be stored in memory. A product like **Ls** can be calculated row by row. That is, we first calculate t_1 for the first ray path, then t_2, etc. A product like $\mathbf{L}^T\mathbf{t} = \mathbf{b}$ can be handled in the same way, although then b_j is not calculated in the jth step, but slowly growing until all rays passing cell j have been treated. Consequently, we never calculate $\mathbf{L}^T\mathbf{L}$ explicitly. This avoids the problem that, although **L** is a sparse matrix with about $NM^{1/D}$ nonzero elements, $\mathbf{L}^T\mathbf{L}$ has a good chance of being a dense matrix [3].

Parts of the model may have been illuminated by only a few rays, or some cells may not have been visited by any ray paths at all. We shall wish to handle indeterminacies in the model by minimizing or restraining

$$\int_R s(\mathbf{r})^2 \, d^D\mathbf{r} = \sum_{k=1}^{M} s_k^2. \tag{9}$$

This solution can be found by means of the singular value decomposition (SVD) of the matrix L. The SVD solution is given by

$$\langle \mathbf{s} \rangle = \mathbf{V}\mathbf{E}^{-2}\mathbf{V}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{t} \tag{10a}$$

where $\mathbf{E}^2$ is a $K \times K$ diagonal matrix of the largest eigenvalues $e_1 \cdots e_k$ of $\mathbf{L}^{\mathrm{T}}\mathbf{L}$, $\mathbf{V}$ the $M \times K$ matrix of the corresponding eigenvectors. Within the K-dimensional subspace spanned by the columns of $\mathbf{V}$, $\langle \mathbf{s} \rangle$ is the unique solution to (8) with the smallest Euclidean norm. It is related to all other solutions of (8), hence to the true solution as well, through the projection matrix $\mathbf{V}\mathbf{V}^{\mathrm{T}}$:

$$\langle \mathbf{s} \rangle = \mathbf{V}\mathbf{V}^{\mathrm{T}}\mathbf{s}. \tag{11}$$

Wiggins [18] proposes to adopt $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ as the "resolution matrix." The idea is that $\mathbf{V}\mathbf{V}^{\mathrm{T}}$ acts as a blurring window, through which we can view the true Earth. However, we should take extreme care with this: the row elements of the resolution matrix do not sum to 1, and (11) does not represent a true physical averaging. In Section 5 of this paper I shall give a more satisfactory approach to resolving power estimation in discretized systems. We note that

$$\mathrm{Var}\{\langle s_i \rangle\} = \sum_{j=1}^{K} V_{ij}^2 e_j^{\ 2}. \tag{12}$$

Since errors are inversely proportional to the magnitude of the eigenvalues, it is advantageous to filter them out in some way. The DLS solution can be constructed by taking, instead of (10),

$$\langle \mathbf{s} \rangle = \mathbf{V}(\mathbf{E}^2 + c^2\mathbf{I})^{-1}\,\mathbf{V}^{\mathrm{T}}\mathbf{L}^{\mathrm{T}}\mathbf{t}. \tag{10b}$$

SVD has so far been the most widely used inversion technique in global seismic tomography [1, 14].

3. Damped Least-Squares Methods

It is possible to decompose the matrix $\mathbf{L}$ in $k+1$ steps into a $(k+1) \times k$ bidiagonal matrix $\mathbf{B}_k$ and matrices $\mathbf{U}_k$ and $\mathbf{V}_k$, whose columns would be orthogonal in the absence of round-off errors [6, 16, 11]:

$$\begin{aligned} \mathbf{L}\mathbf{V}_k &= \mathbf{U}_{k+1}\mathbf{B}_k \\ \mathbf{L}^{\mathrm{T}}\mathbf{U}_{k+1} &= \mathbf{V}_k\mathbf{B}_k^{\mathrm{T}} + \mathbf{r}_k\mathbf{e}_{k+1}^{\mathrm{T}} \end{aligned} \tag{13}$$

where $\mathbf{r}_k$ is a vector of residuals that (at least in theory) goes to zero in M or less iterations in the absence of round-off errors. The recursion involves only multiplications with $\mathbf{L}$ and $\mathbf{L}^{\mathrm{T}}$. Its storage requirements are modest. Only the last

columns of $\mathbf{U}$ and $\mathbf{V}$ are needed in memory. Together with the $2k$ elements of $\mathbf{B}_k$ the algorithm requires a minimum storage of $N+M+2k$, apart from possible storage of elements of $\mathbf{L}$.

This algorithm can be exploited in several ways. First of all, the largest eigenvalues of $\mathbf{T}=\mathbf{B}^T\mathbf{B}$ converge to those of $\mathbf{L}^T\mathbf{L}$, and can be calculated conveniently with the use of Sturm sequences [6, 13]. The effect of round-off errors is to introduce artificial multiplicities of single eigenvalues. These artificial eigenvalues can easily be identified by monitoring the convergence as k increases [8].

For SVD, the eigenvectors must be computed as well. Unfortunately this involves transformation matrices of order $M \times k$. Thus, use of the algorithm in this way is expected to be limited to systems of low rank, which can be described with not too many eigenvectors. Nevertheless it may be used to apply SVD to tomographic systems of somewhat larger size than currently in vogue, retaining all the advantages of SVD.

Alternatively, Paige and Saunders [11] have employed (13) for a stable, recursive DLS algorithm. In their method, the kth approximation to the solution $\mathbf{s}$ is defined to be $\mathbf{s}_k=\mathbf{V}_k\,\mathbf{y}_k$, where $\mathbf{y}_k$ is the DLS solution to the $(k+1)\times k$ system $\mathbf{B}_k\,\mathbf{y}_k=\mathbf{e}_1\beta/|\mathbf{t}|$. The method is closely related to the conjugate gradient method, in which the residual vectors of successive iterations are mutually orthogonal, and the corrections applied to $\mathbf{s}_k$ in each iteration are mutually conjugate with respect to the matrix $\mathbf{L}$.

Because of the orthogonalization involved, the algorithm should theoretically converge in N steps or less. Round-off errors do make trouble, however. Nevertheless one may expect that the convergence rate of such methods is faster than that of simpler iterative methods. As an example of such methods we chose the algorithm proposed by Dines and Lytle [5]. It is a stationary iterative process, i.e., the correction to the approximate solution $\mathbf{s}^{(k)}$ after k steps is chosen proportional in magnitude to the residual vector $\mathbf{r}^{(k)}=\mathbf{t}-\mathbf{L}\mathbf{s}^{(k)}$:

$$\mathbf{s}^{(k)}=\mathbf{s}^{(k-1)}+\mathbf{H}\mathbf{r}^{(k-1)}. \tag{14}$$

Since the residual $\mathbf{r}^{(k)}$ is recalculated after each step, round-off errors are usually not important in stationary iterative processes. In the Dines-Lytle method one chooses

$$\mathbf{H}=\mathbf{S}\mathbf{L}^T\mathbf{D} \tag{15}$$

where $\mathbf{D}=\mathrm{diag}(D_{ii})$ with $D_{ii}=\{\sum_{j=1}^{M} L_{ij}^2\}^{-1}$ and $\mathbf{S}=\mathrm{diag}(n_i)$, n_i being the number of nonzero elements in the ith column of $\mathbf{L}$.

The reason for choosing this algorithm is that its convergence to the minimum norm least-squares solution is assured [7]. When stopped with a limited number of iterations completed, the contributions of small eigenvalues are damped out with a filter only slightly different from that in (10b). This sets the Dines-Lytle algorithm apart from other popular row-action methods such as the algebraic reconstruction technique, which do converge to a minimum norm, but not necessarily to the least-squares solution.

4. Numerical Experiments

The Perfect Data Case

In order to test the numerical performance of the algorithms, I devised a simple experiment. The model consists of a rectangle of $20 \times 20 = 200$ cells. Although the physical dimensions are irrelevant to the results, this actually models part of the upper mantle (width 600 km, depth 300 km). As in global seismology, the model is illuminated from below by seismic rays from a source located far away (Fig. 1). Each of the 20 surface cells is covered by a seismograph.

The velocities in the model differ by up to 1% from a standard model. Velocity deviations decrease gradually from the upper left to the lower right corner of the model. There is, however, a sharp discontinuity with a sudden increase in velocity, tilted at about 45 degrees, and a sharply defined low-velocity body 5 cells wide and 2 cells deep. Although by itself not realistic, the model combines features commonly encountered in the Earth, such as a high-velocity slab, a gradual velocity change in lateral direction, and a (magma) body of low velocity.

We shall study two different types of illumination. In the "wide-angle" case, 20 rays arrive in a regular fan with angles between -35 and $+5$ degrees to the vertical. In the "narrow-angle" case, the 20 rays arrive in a narrow fan between -5 and -1 degrees to the vertical. Thus, there are 400 data points, 200 unknows.

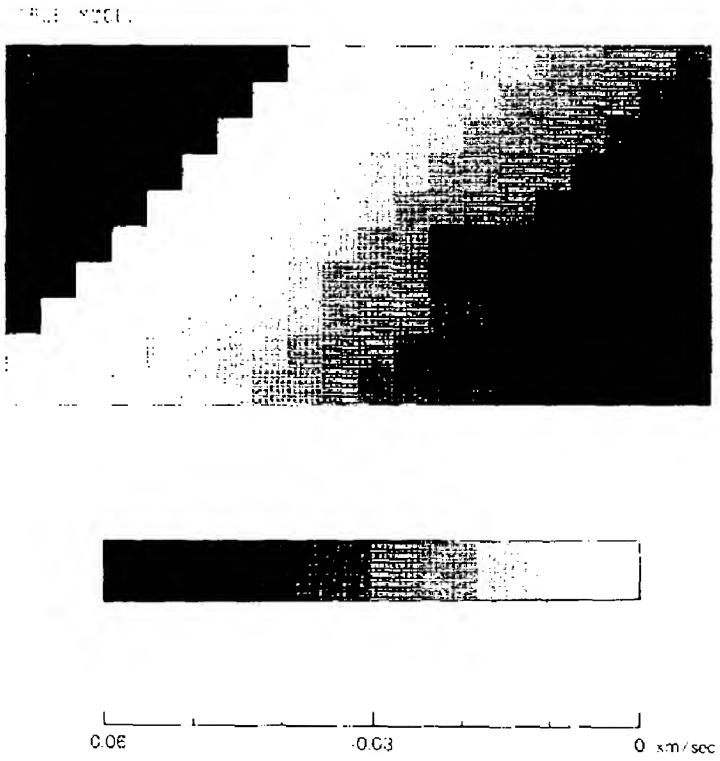

Fig. 1. The test model used in the calculations. The darkness in each of the 10×20 cells indicates the drop in velocity with respect to a homogeneous model. The colour scale is indicated below.

The eigenvalues of the least-squares system (8) have been calculated with the Lanczos algorithm, and one may study the condition of the system by looking at the eigenvalues of $\mathbf{L}^T\mathbf{L}$ (Fig. 2). For the narrow-angle illumination, there are essentialy 20 large eigenvalues, after which the magnitude of the eigenvalues decreases rapidly to zero. Obviously, the average velocity in each of the 20 columns in Fig. 1 is well-resolved, but any detail in the vertical direction can only be obtained with a large standard deviation in the solution. The decrease in the magnitude of the eigenvalues for the wide-angle case is much more gradual. The eigenvalues still have a considerable magnitude when eigenvalues of the narrow-angle case have all but disappeared. This reflects the larger variation in illumination angles as well as the improved vertical resolution.

For both cases the eigenvalues were calculated with 200 iterations. In the wide-angle case 54 eigenvalues converged, and in the narrow-angle case only 48. It is evident that many iterations are necessary, more than the number of eigenvalues

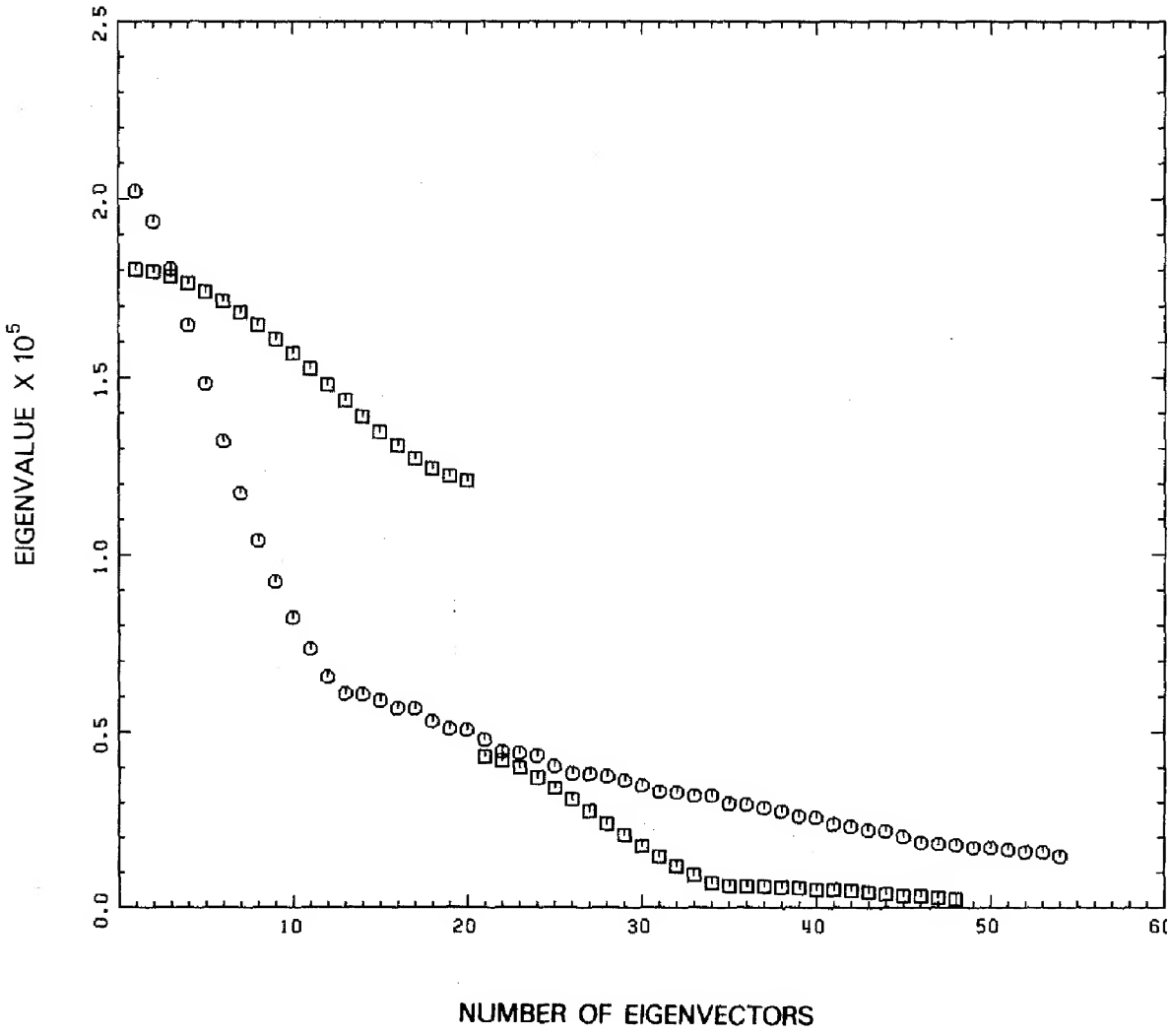

FIG. 2. The eigenvalue patterns for wide-angle (○) and narrow-angle (□) illumination, using Paige–Lanczos algorithm. Eigenvalues are in order of decreasing magnitude.

obtained. This is especially so when small eigenvalues, usually much closer to each other, are desired.

Because the usefulness of SVD in large systems relies heavily on the possibility of truncating the system at a very low value of the effective rank, I have investigated how well the system of equations (7) can be fitted with only K eigenvectors, K being small with respect to the number of cells. To quantify the fit we introduce the normalized misfit R of a solution $\mathbf{s}$ to (7) as

$$R=\left\{\sum_{i,j}(L_{ij}s_j-t_i)^2\right\}^{1/2}\left\{\sum_j t_j^2\right\}^{-1/2} \tag{16}$$

For both the wide-angle and the narrow-angle illumination the results are shown in Fig. 3 for $1<K<40$. The solutions for $K=40$ are shown in Fig. 4. When compared to Fig. 1, we see that the resolution for the narrow-angle case is quite dis-

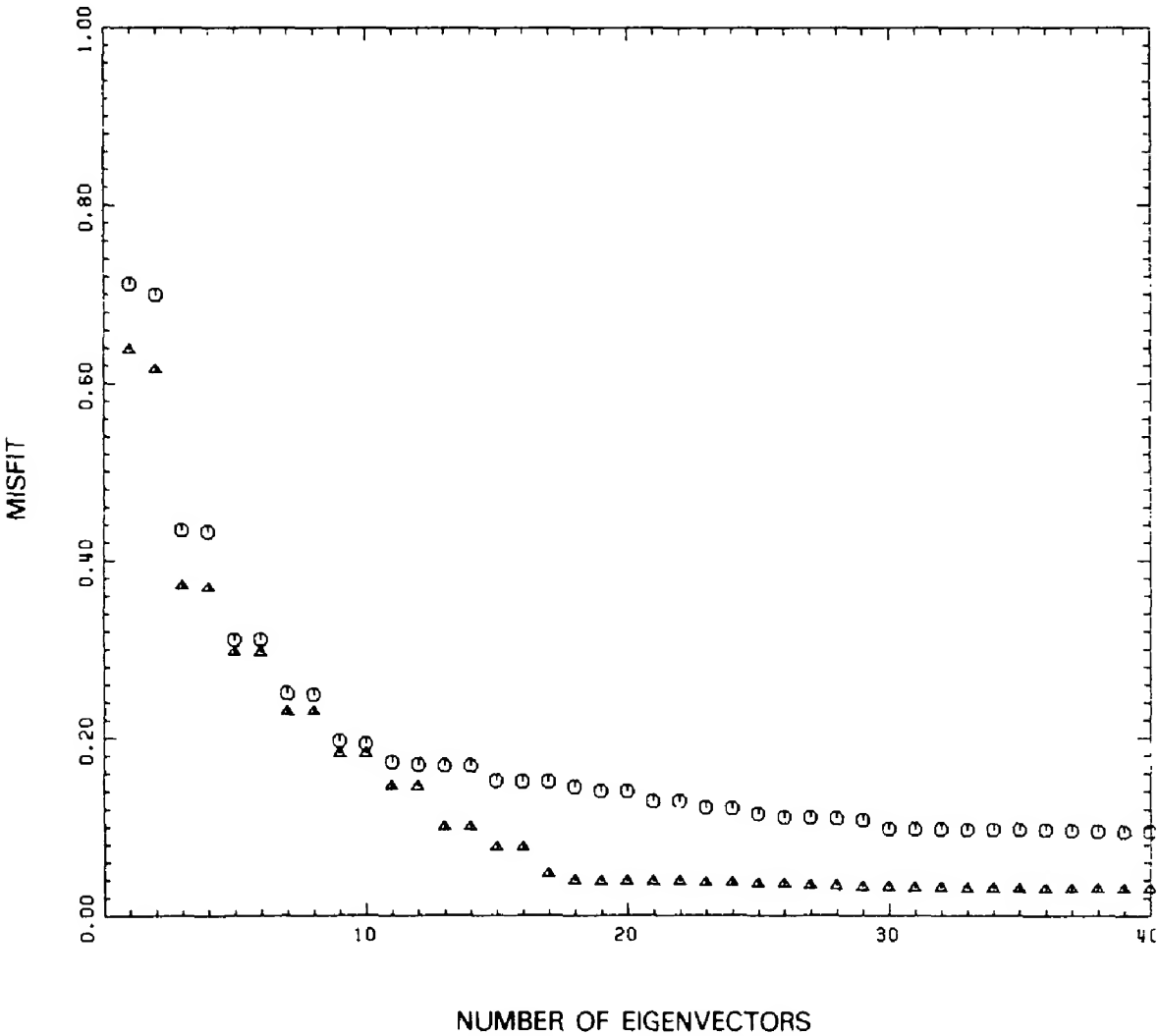

FIG. 3. Misfit of the solutions obtained with singular value decomposition, using only a few eigenvectors belonging to the largest eigenvalues and using Paige–Lanczos algorithm; wide-angle (○) and narrow-angle (△) illumination.

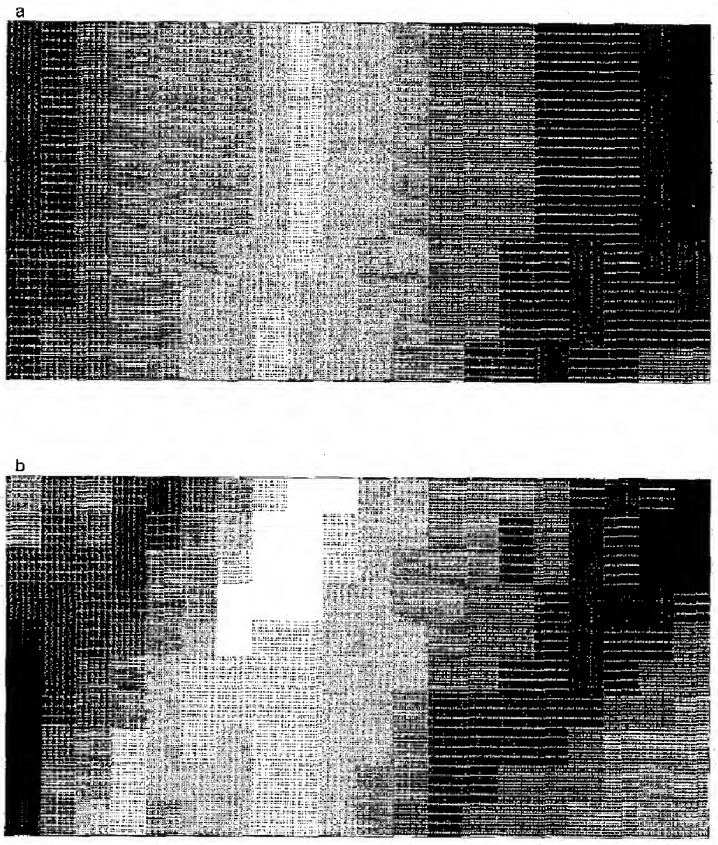

FIG. 4. SVD solution for 40 eigenvectors. (a) Narrow-angle illumination, normalized misfit 0.030, and (b) wide-angle illumination, normalized misfit 0.095.

appointing, as was expected from the eigenvalues. For the wide-angle illumination we see that the general trend is recovered with 40 eigenvectors, but the sharp transition to higher velocities is quite blurred and the magma body is not recognizable. In the following tests we shall concentrate on the wide-angle illumination.

With less computing effort the Paige–Saunders method performs far better (Fig. 5a). In only 20 iterations the misfit R is down to 0.007, as compared to 0.095 for the SVD method. Most of the results in this paper were obtained with an early coded version of the algorithm, which did not incorporate damping. A later version, published as algorithm LSQR [12], does include damping and is optimized with respect to memory usage. Whenever results on computational speed are reported in this paper, it is for the undamped version of the Paige–Saunders algorithm. The differences with the damped version are not important, however.

The computational effort to do one iteration is about equal for the two algorithms, but the SVD requires eigenvector determination as well. It took 400 iterations to solve (8) to close to machine precision (11 digits) as shown in Fig. 5b. For the narrow-angle case a comparable fit can easily be obtained, but the resulting model does not show the sharp contrasts of the original in Fig. 1. This again shows that the narrow-angle case is truly singular at the level of precision used.

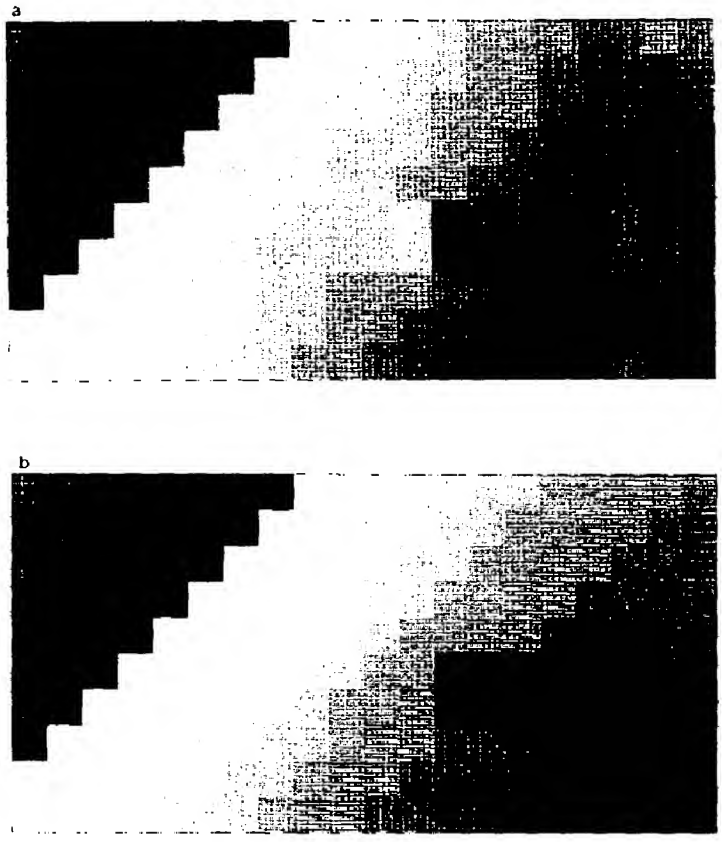

FIG. 5. Paige-Saunders solution for wide-angle illumination; (a) undamped solution after 20 iterations, misfit 0.007, and (b) after convergence to almost machine precision in 400 iterations.

Finally, we used the Dines-Lytle procedure to solve for the same data set (Fig. 6). The models obtained after 20 and 400 iterations look quite reasonable. The sharp contrast at the high-velocity slab is resolved after 400 iterations, but not the magma body, although its left boundary is just becoming visible. We know that Dines–Lytle should eventually converge to the right solution, but apparently there is still a lot of information contained in the small eigenvalues.

In terms of computation time, the Paige-Saunders algorithm was found to be superior to Dines–Lytle when we consider the misfit R (Fig. 7). An initial lead of the Dines–Lytle algorithm, originating from the choice of the starting vector, is soon regained. At a misfit level of 5%, Paige-Saunders is about 5 times faster in the wide-angle case, and at the 1% level it is 10 times faster in both experiments.

4b *The Noisy Data Case*

The same algorithms have been compared for their filtering properties on noise affecting the data. To this end I added randomly oriented vectors to the data vector $\mathbf{t}$ for the case of wise-angle illumination. The length of the random vector was fixed at 50% of $|\mathbf{t}|$ to simulate a set of data that is badly degraded by observation errors. The results are shown in Fig. 8.

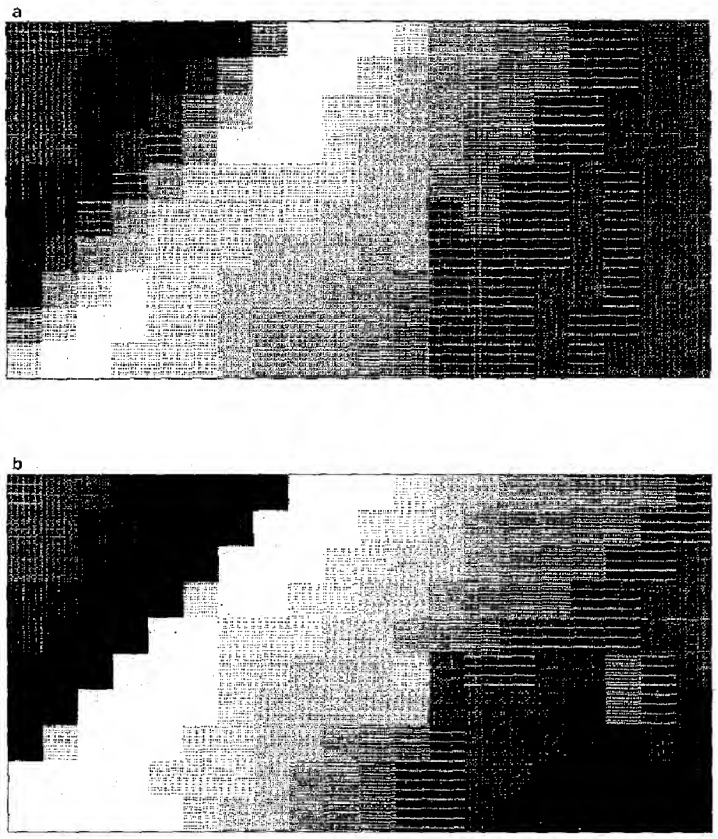

FIG. 6. Dines–Lytle solutions for wide-angle illumination (a) after 20 iterations, misfit 0.031, and (b) after 400 iterations, misfit 0.005.

The SVD solution (8a) is quite successful in filtering out the velocity trend, although the velocity contrast at the fault is smoothed and the low velocities of the magma body are smeared out into the dominant ray direction. The undamped Paige–Saunders, (8b) stopped after five iterations, gives a much more erratic solution, as does the Dines–Lytle after 20 iterations (8c).

The introduction of damping into the Paige–Saunders algorithm did improve the resolution of the velocity trend (Fig. 9). The introduction of damping into the mathematical solution is equivalent to introducing hypothetical ray segments in each cell separately, with zero delay time, *after* the system has been scaled to unit variance in the data. With some experimentation we found that a reasonable solution was obtained for a damping equivalent to a ray length of 10 km in each cell *before* scaling, i.e., about 1.6% of the average ray path length in each cell.

In a separate test, run with a 5% random error vector, all three methods continued to give acceptable solutions. In this case the undamped Paige–Saunders algorithm started to give erratic solutions when more than about 15 iterations were calculated.

For large errors, the damped Paige–Saunders algorithm is to be preferred. It is

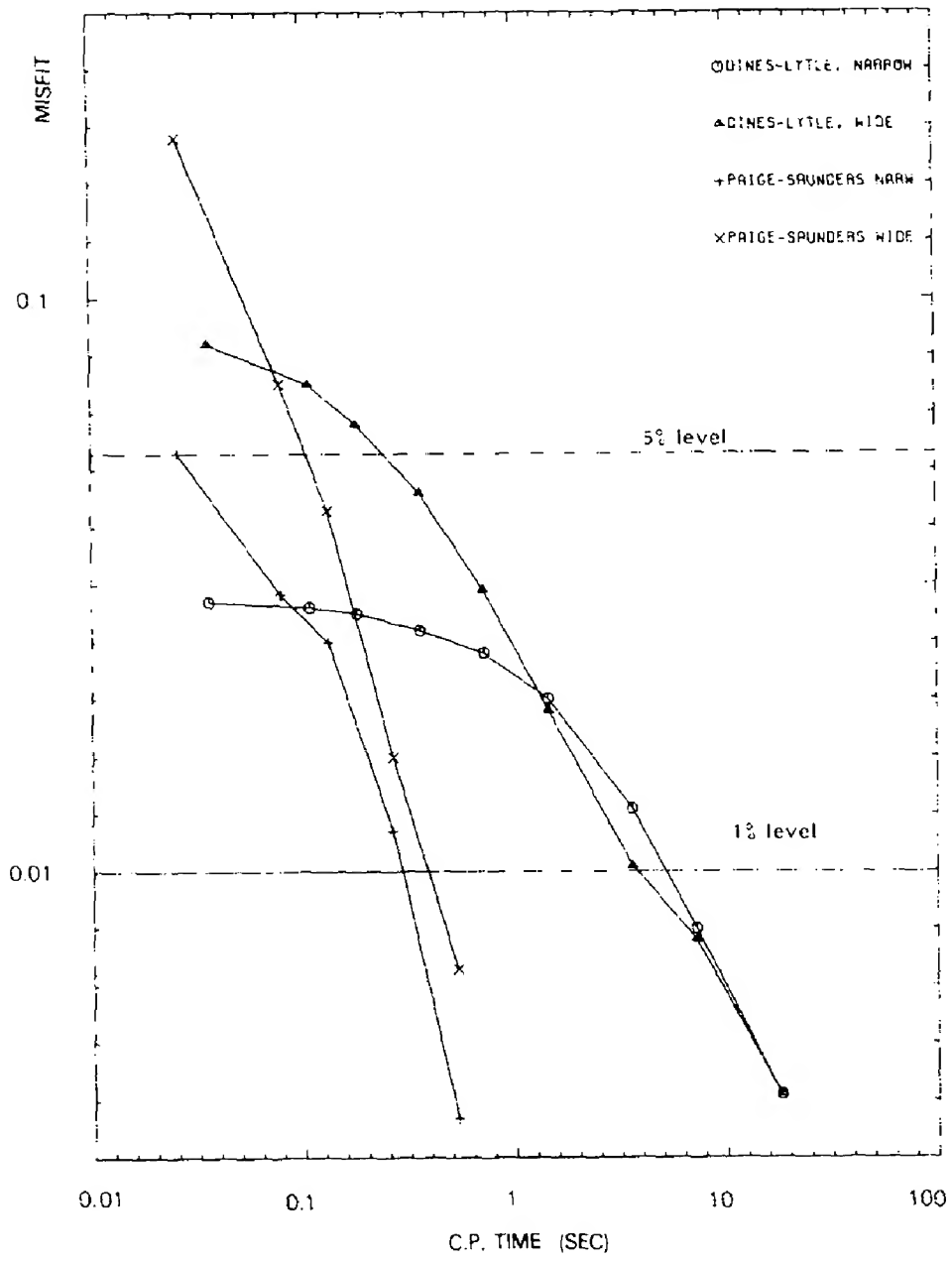

FIG. 7. Decrease of the normalized length of the residual vector as a function of central processor time on a CDC Cyber 175.

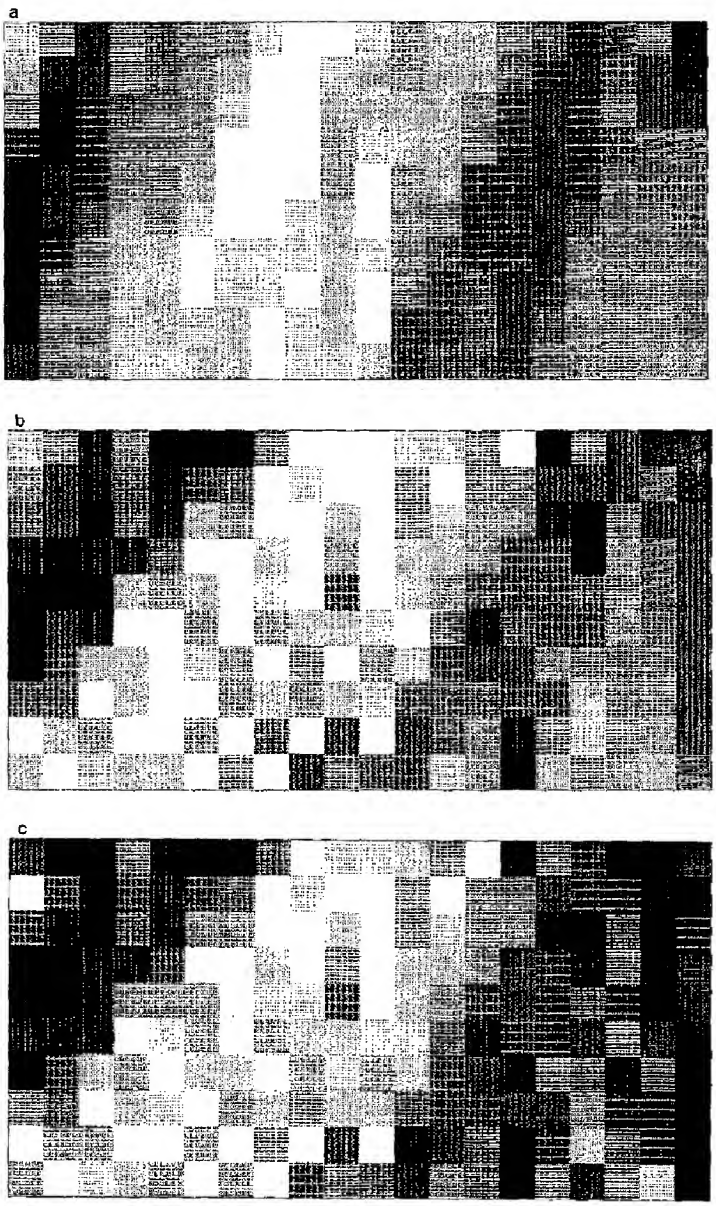

FIG. 8. The influence of 50% noise on the solutions for wide-angle illumination. (a) SVD for 40 eigenvectors, misfit 0.284, (b) undampened Paige–Saunders solution after 5 iterations, misfit 0.254, and (c) Dines–Lytle after 20 iterations, misfit 0.237.

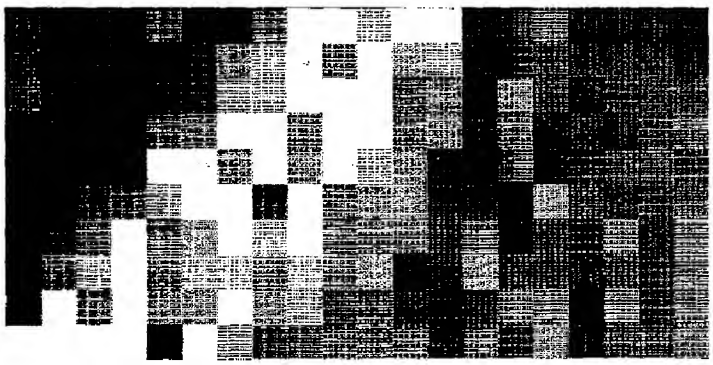

FIG. 9. Damped Paige–Saunders solution in the 50% noise case, misfit 0.337.

superior in convergence speed to Dines-Lytle, and requires much less back-up storage than SVD, since no eigenvectors are needed. However, it is evident from Figs. 8 and 9 that many unreal phenomena may turn up in the solution. A separate analysis is needed to establish what resolution is warranted by the data and what is not. This problem will be attacked in the next section.

5. Resolution Analysis

For the description of the resolution we use the concept of an averaging kernel, originally formulated by Backus and Gilbert [2] for continuous models. Their analysis needs to be modified for discrete, parameterized models.

The basic idea in Backus–Gilbert resolution theory is to estimate a local average of the model around $\mathbf{r}_0$. This average $\langle s(\mathbf{r}_0)\rangle$ is a linear function of the data

$$\langle s(\mathbf{r}_0)\rangle = \sum_j b_j(\mathbf{r}_0)\, t_j = \sum_{j,k} b_j(\mathbf{r}_0)\, L_{jk} s_k. \tag{17}$$

Our task is to determine $b_j(\mathbf{r}_0)$. A local average of the real Earth can be constructed with a peak-shaped function $A(\mathbf{r}, \mathbf{r}_0)$:

$$\langle s(\mathbf{r}_0)\rangle = \int_R A(\mathbf{r}, \mathbf{r}_0)\, s(\mathbf{r})\, d^D\mathbf{r} \tag{18}$$

where

$$\int_R A(\mathbf{r}, \mathbf{r}_0)\, d^D\mathbf{r} = 1. \tag{19}$$

Ideally, the amplitude of $A(\mathbf{r}, \mathbf{r}_0)$ is negligible outside a small region round $\mathbf{r}_0$. As a measure of the width of A around $\mathbf{r}_0$ we use

$$W(\mathbf{r}_0) = c_D \int_R A(\mathbf{r}, \mathbf{r}_0)^2\, |\mathbf{r} - \mathbf{r}_0|^{D+1}\, d^D\mathbf{r} \tag{20}$$

where the factor $|\mathbf{r} - \mathbf{r}_0|^{D+1}$ ensures the correct dimensionality of W, and c_D is a scaling factor gauged with help of a simple form of A. For example, if A is 0 outside, and constant inside a region with radius $\rho/2$, c_D ensures that $W(\mathbf{r}_0) = \rho$ for $c_1 = 12$, $c_2 = 5\pi$, $c_3 = 56\pi/9$, respectively. If we develop

$$A(\mathbf{r}, \mathbf{r}_0) = \sum_{i=1}^{M} a_i(\mathbf{r}_0)\, h_i(\mathbf{r}) \tag{21}$$

we find from (18)

$$\langle s(\mathbf{r}_0)\rangle = \sum_{k=1}^{M} a_k(\mathbf{r}_0)\, s_k \tag{22}$$

so that we identify from (17)

$$a_k(\mathbf{r}_0) = \sum_{j=1}^{N} L_{jk} b_j(\mathbf{r}_0). \tag{23}$$

The relationship between $W(\mathbf{r}_0)$ and $b_j(\mathbf{r}_0)$ is found by substitution of (4), (21), and (23) into (20):

$$W(\mathbf{r}_0) = \sum_{i,j,k} f_i L_{ji} L_{ki} b_j(\mathbf{r}_0)\, b_k(\mathbf{r}_0) \tag{24}$$

where

$$f_i = c_D \int_R |\mathbf{r} - \mathbf{r}_0|^{D+1} h_i(\mathbf{r})^2 \, d^D\mathbf{r} \approx d_i^2 \tag{25}$$

with $d_i^2 = |\mathbf{r}_i - \mathbf{r}_0|^{D+1} c_D$ and $\mathbf{r}_i$ the center of gravity of cell i. The error in the average $\langle s(\mathbf{r}_0)\rangle$ is also a quadratic function of the $b_j(\mathbf{r}_0)$:

$$\begin{aligned} \varepsilon^2(\mathbf{r}_0) &= \mathrm{Var}\langle s(\mathbf{r}_0)\rangle \\ &= \sum_{i,j=1}^{N} b_i(\mathbf{r}_0)\, b_j(\mathbf{r}_0)\, \mathrm{Cov}(t_i, t_j) = \sum_{j=1}^{N} b_j(\mathbf{r}_0)^2. \end{aligned} \tag{26}$$

We are now ready to determine the factors $b_j(\mathbf{r}_0)$. Facing a choice between minimizing $W(\mathbf{r}_0)$ or $\varepsilon^2(\mathbf{r}_0)$ we introduce a trade-off parameter w $(0 \leqslant w \leqslant \infty)$ and minimize

$$J(w, \mathbf{r}_0) = W(\mathbf{r}_0) + w^2 \varepsilon^2(\mathbf{r}_0) \tag{27}$$

under the constraint (19), or

$$\sum_{i,j} v_i^{1/2} L_{ji} b_j(\mathbf{r}_0) = 1. \tag{28}$$

If we use the approximation for f_i in (24) and set

$$P_{ij} = d_j L_{ij} \qquad c_j = \sum_i L_{ji} v_i^{1/2} \tag{29}$$

then we minimize

$$J = \mathbf{b}^{\mathrm{T}}[\mathbf{P}\mathbf{P}^{\mathrm{T}} + w^2\mathbf{I}]\,\mathbf{b} \qquad \text{with} \quad \mathbf{c}^{\mathrm{T}}\mathbf{b} = 1. \tag{30}$$

Since P is a sparse matrix, the solution of this system poses no extra storage problems. Equation (30) could be solved by the introduction of a Lagrange multiplier, but in our experience the system is in this case very ill-conditioned and the LSQR algorithm failed to converge in a reasonable number of iterations. Saunders (personal communication, 1983) recommends expressing b_1 (or any other element

of **b**) first in terms of the others, and substitute in J before minimizing. This has the added convenience of formulating the system directly as a DLS problem, for subsequent solving with the Paige–Saunders algorithm. We truncate **b** and **c** to vectors **p** and **q** of length $N-1$,

$$\mathbf{p} = [b_2, b_3, \ldots, b_N], \qquad \mathbf{q} = c_1^{-1}[c_2, c_3, \ldots, c_N]$$

so that

$$b_1 = c_1^{-1} - \mathbf{q}^T\mathbf{p}$$

$$\mathbf{b} = \mathbf{B}\mathbf{p} + c_1^{-1}\mathbf{e}_1$$

where $\mathbf{e}_1$ is the unit vector $[1, 0, \ldots, 0]$ and

$$\mathbf{B} = \begin{bmatrix} -\mathbf{q}^T \\ \mathbf{I}_{N-1} \end{bmatrix}$$

if we write

$$\mathbf{A} = \begin{bmatrix} \mathbf{P}^T \\ w\mathbf{I}_N \end{bmatrix}.$$

Equation (30) becomes equivalent to

$$\text{Min}|\mathbf{p}^T\mathbf{B}^T(\mathbf{A}^T\mathbf{A})\,\mathbf{B}\mathbf{p} + 2\mathbf{p}^T\mathbf{B}^T(\mathbf{A}^T\mathbf{A})\,c_1^{-1}\mathbf{e}_1|$$

or

$$\mathbf{B}^T\mathbf{A}^T\mathbf{A}\mathbf{B}\mathbf{p} + \mathbf{B}^T\mathbf{A}^T\mathbf{A}c_1^{-1}\mathbf{e}_1 = 0.$$

These are the normal equations belonging to another minimization problem

$$\text{Min}\,|\mathbf{A}\mathbf{B}\mathbf{p} + c_1^{-1}\mathbf{A}\mathbf{e}_1|.$$

Substituting for **A** and **B** we find that this is essentially a DLS problem with damping parameter w:

$$\mathbf{A}\mathbf{B}\mathbf{p} = \begin{pmatrix} \mathbf{P}^T\mathbf{B} \\ \hline -w\mathbf{q}^T \\ \hline w\mathbf{I}_{N-1} \end{pmatrix}\mathbf{p} = -c_1^{-1}\begin{pmatrix} \mathbf{p}^T\mathbf{e}_1 \\ \hline w \\ \hline 0 \end{pmatrix}. \tag{31}$$

The system (31) has been solved without difficulty. w governs the trade-off between kernel width and the variance of the error.

An example of an averaging kernel, computed for the resolution of the center element in the model, is shown in Fig. 10. The error vector here is 5%. This corresponds to a standard deviation in measured seismic delay times of about 0.1 s.

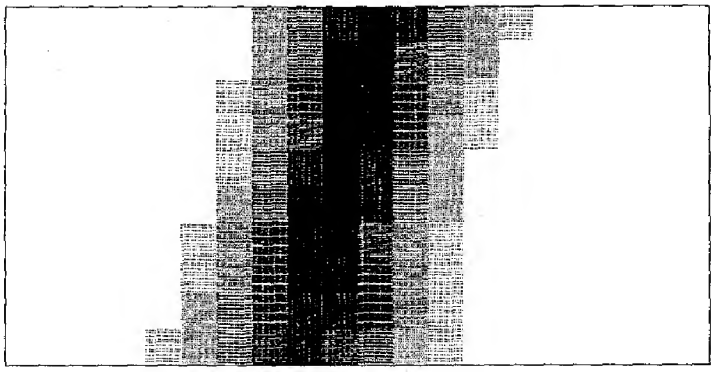

FIG. 10. Shape of the averaging kernel $A(\mathbf{r}, \mathbf{r}_0)$ where $\mathbf{r}_0$ is at the center of the rectangular model for the case of 5% noise, wide-angle illumination. The standard deviation in the average is 0.001 km/*s*.

which is good but not unrealistic. The elongated shape of the kernel, with a length comparable to the height of the model, but a width of only 3 or 4 cells, reflects the preferred orientation of the rays crossing the model.

The trade-off curve in Fig. 11 shows how the width is a function of the standard deviation in the local average. There is apparently little choice: the preferred ray orientation leads to very precise averages (0.001 km/s), with large location uncertainty (220 km). Efforts to exchange the uncertainty are futile: at best the averaging length goes down to 200 km. In fact, the simple shape of the trade-off curve suggests that one may obtain a reasonably good approximation to it by performing only two calculations, one on each end of the curve.

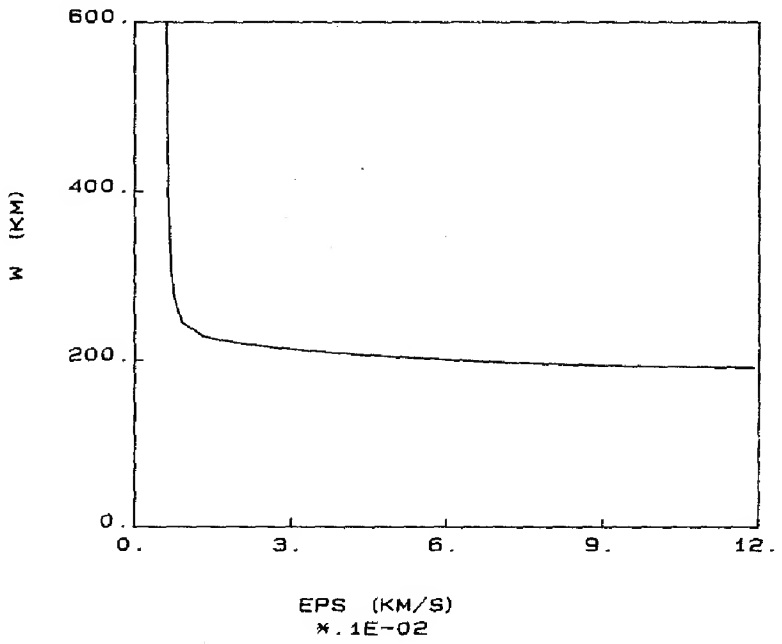

FIG. 11. Trade-off between averaging length W and the standard deviation in the average for the center element of the model, for the case of 5% noise, wide-angle illumination.

It should be noted that we have assumed our data to be unbiased, i.e., the mean error to be 0. If this is not the case, the solution will be biased as well, and the calculation of the error with (26) will be optimistic. Moreover, if our assumption that the data are uncorrelated does not hold, the resolution kernels themselves will be in error. This situation is difficult to handle. Even if we have an estimate of the data correlations, scaling the system to unit variance will destroy the sparsity of $\mathbf{L}$ in (7) and greatly increase the computational effort needed. If $\mathbf{C}_d$ is not diagonal, the best strategy is probably to accumulate groups of strongly correlated data into one equation by summing.

6. Conclusions

In test calculations, the Paige-Saunders LSQR algorithm proved to be superior to the Lanczos SVD and Dines Lytle methods for the calculations of DLS models in seismic tomography. A feasible method to calculate the resolution, using the LSQR algorithm, was developed. Although the computation time for the resolution of one spot in the model is comparable to that needed to calculate the solution itself, resolution calculations are badly needed to screen out the effects of observation errors.

Test calculations for the resolution in a realistic tomographic experiment show that it will be very difficult to obtain sufficient resolution in the upper mantle, where rays from teleseismic events arrive in narrow bundles, not far deflected from the vertical. An exception to this is in regions of deep seismicity.

Acknowledgments

Discussions with Dr. Henk van der Vorst convinced me of the need to investigate different iterative methods for their performance on tomography with noisy seismic data. Professor Jon Claerbout made a short stay at Stanford possible, where I enjoyed discussions with him and his students, and where Dr. Michael Saunders helped me quickly out of a nasty numerical stability problem that turned up in the resolving power calculations.

References

1. L. Aki, A. Christofferson, and E. Husebye, *Bull. Seismol. Soc. Amer.* **66** (1976), 501.
2. G. Backus and F. Gilbert, *Geophys. J. Astronom. Soc.* **13** (1967), 247.
3. A. Björck and T. Elfving, *BIT* **19** (1979), 145.
4. R. W. Clayton and R. P. Comer, *EOS Trans. Amer. Geophys. Union* **64** (1983), 776.
5. K. A. Dines and R. J. Lytle, *Proc. IEEE* **67** (1979), 1065.
6. G. Golub and W. Kahan, *Numer. Anal. J. SIAM B* **2** (1965), 205.
7. S. Ivansson, *Geophys. J. Astronom. Soc.* **75** (1983), 855.

8. J. M. VAN KATS AND H. A. VAN DER VORST, "Automatic Monitoring of Lanczos Schemes for Symmetric or Skew-Symmetric Generalized Eigenvalue Problem," Technical Report TR-7, ACCU, University of Utrecht, Netherlands, 1977.
9. G. A. MCMECHAN, *Geophys. J. R. Astronom. Soc.* **74** (1983), 601.
10. G. NOLET, "The Solution of the Inverse Problem in Geophysical Interpretation" (R. Cassinis, Ed.), p. 9, Plenum, Nex York, 1981.
11. C. C. PAIGE AND M. A. SAUNDERS, *ACM Trans. Math. Software* **8** (1982), 43.
12. C. C. PAIGE AND M. A. SAUNDERS, *ACM Trans. Math. Software* **8** (1982), 195.
13. G. PETERS AND J. H. WILKINSON, *in* "Linear Algebra" (J. H. Wilkinson and C. Reinsch, Eds.), Springer-Verlag, New York/Berlin, 1971.
14. C. SPENCER AND D. GUBBINS, *Geophys. J. R. Astronom. Soc.* **63** (1980), 95.
15. C. H. THURBER AND W. L. ELLSWORTH, *Bull. Seismol. Soc. Amer.* **70** (1980), 1137.
16. J. VAN HEIJST, JACOBS, AND J. SCHERDERS, "Kleinste kwadraten problemen," MSc. thesis, Dept. Mathematics, Eindhoven Polytechnic, Eindhoven, Netherlands, 1976.
17. D. N. WHITCOMBE, *Geophys. J. R. Astronom. Soc.* **69** (1982), 635.
18. R. A. WIGGINS, *Rev. Geophys. Space Phys.* **10** (1972), 251.